WebScrape

DevOps

2de fase IT Factory

Jasper Vermissen

Academiejaar 2021-2022

Campus Geel, Kleinhoefstraat 4, BE-2440 Geel

THOMAS MORE

1 WAT EN HOE

Ik heb voor mijn project van DevOps de keuzen gemaakt om meer te leren over webscraper. Ik ben begonnen met de gegeven links te bekijken. Ik vond wel wat info maar vergeleken met de filmpjes dat ik gevonden hebt op youtube is er groot verschil. Een groot deel van mijn info komt van de youtuber Scrapax. En met deze info heb ik mijn project afgelegd.

2 CODE

2.1 Youtube

```csharp
using OpenQA.Selenium.Chrome;
using OpenQA.Selenium;
using System;
using System.Collections.Generic;
using System.Linq;
using System.Text;
using System.Threading.Tasks;
using System.Diagnostics;
using static System.Net.WebRequestMethods;
using System.Collections.ObjectModel;
using Pipelines.Sockets.Unofficial.Buffers;
using System.IO;



namespace webscraping
{
    class Program
    {
        private static IWebDriver driver;
        static void Main(string[] args)
        {
            String Search = "mrbeast";
            String url = "https://www.youtube.com/results?search_query=" +
Search;
            int vcount = 1;
            driver = new ChromeDriver();
            driver.Navigate().GoToUrl(url);
            var discoverButton =
driver.FindElement(By.XPath("//*[@id=\"content\"]/div[2]/div[6]/div[1]/ytd-
button-renderer[2]/yt-button-shape"));
            discoverButton.Click();
            Thread.Sleep(1000);
            By elem_videos_link = By.Id("dismissible");
            ReadOnlyCollection<IWebElement> videos =
driver.FindElements(elem_videos_link);
            Console.WriteLine("Total number of videos in " + url + " are " +
videos.Count);



            /* Go through the Videos List and scrap the same to get the
attributes of the videos in the channel */
            foreach (IWebElement video in videos)
            {
                if (vcount <= 5)
                {
                    string str_title, str_views, str_name;
                    IWebElement elem_video_title =
video.FindElement(By.Id("video-title"));
                    str_title = elem_video_title.Text;

                    IWebElement elem_video_views =
video.FindElement(By.Id("metadata-line"));
                    str_views = elem_video_views.Text;

                    IWebElement elem_video_name =
video.FindElement(By.Id("channel-info"));
                    str_name = elem_video_name.Text;
```

```
                    Console.WriteLine("****** Video " + vcount + " ******");
                    Console.WriteLine("Video Title: " + str_title);
                    Console.WriteLine("Video Views: " + str_views + "
geupload");
                    Console.WriteLine("Channel name: " + str_name);
                    Console.WriteLine("\n");



                }

                vcount++;
            }

        }
    }
}
```

2.2　Job

```csharp
using OpenQA.Selenium.Chrome;
using OpenQA.Selenium;
using System;
using System.Collections.Generic;
using System.Linq;
using System.Text;
using System.Threading.Tasks;
using System.Diagnostics;
using static System.Net.WebRequestMethods;
using System.Collections.ObjectModel;

namespace webscraping
{
    class Program
    {
        private static IWebDriver driver;
        static void Main(string[] args)
        {

            String url = "https://www.ictjob.be/nl/";
            int vcount = 1;
            driver = new ChromeDriver();
            driver.Navigate().GoToUrl(url);
            var element = driver.FindElement(By.XPath("//*[@id=\"keywords-
input\"]"));
            element.SendKeys("IT");
            element.Submit();
            By elem_job_link = By.ClassName("job-info");
            ReadOnlyCollection<IWebElement> jobs =
driver.FindElements(elem_job_link);
            Console.WriteLine("Total number of jobs in " + url + " are " +
jobs.Count);



            foreach (IWebElement job in jobs)
            {
                if (vcount <= 5)
                {
                    string str_title, str_views, str_rel, str_link;
```

```
                          IWebElement elem_job_title =
job.FindElement(By.ClassName("job-title"));
                          str_title = elem_job_title.Text;

                          IWebElement elem_job_name =
job.FindElement(By.ClassName("job-company"));
                          str_views = elem_job_name.Text;

                          IWebElement elem_job_details =
job.FindElement(By.ClassName("job-details"));
                          str_rel = elem_job_details.Text;

                          IWebElement elem_job_links =
job.FindElement(By.ClassName("search-item-link"));
                          str_link = elem_job_links.Text;

                          Console.WriteLine("******* Job " + vcount + " ******");
                          Console.WriteLine("Job Title: " + str_title);
                          Console.WriteLine("Company: " + str_views);
                          Console.WriteLine("Details: " + str_rel);
                          Console.WriteLine("link: " + str_link);
                          Console.WriteLine("\n");

                      }

                      vcount++;
                  }



              }
          }
}
```

2.3   Bol

```csharp
using OpenQA.Selenium.Chrome;
using OpenQA.Selenium;
using System;
using System.Collections.Generic;
using System.Linq;
using System.Text;
using System.Threading.Tasks;
using System.Diagnostics;
using static System.Net.WebRequestMethods;
using System.Collections.ObjectModel;

namespace webscraping
{
    class Program
    {
        private static IWebDriver driver;
        static void Main(string[] args)
        {
            String Searh = "Desktops";
            String url = "https://www.bol.com/nl/nl/s/?searchtext=" + Searh;
            int vcount = 1;
            driver = new ChromeDriver();
            driver.Navigate().GoToUrl(url);
            var discoverButton = driver.FindElement(By.Id("js-first-screen-accept-all-button"));
```

```csharp
            discoverButton.Click();
            By elem_artikelen_link = By.ClassName("product-item--row");
            ReadOnlyCollection<IWebElement> artikelen =
driver.FindElements(elem_artikelen_link);
            Console.WriteLine("Total number of artikelen in " + url + " are " +
artikelen.Count);


            /* Go through the Videos List and scrap the same to get the
attributes of the videos in the channel */
            foreach (IWebElement artikel in artikelen)
            {
                if (vcount <= 5)
                {
                    string str_title, str_views, str_prijs;
                    IWebElement elem_artikel_title =
artikel.FindElement(By.ClassName("product-title"));
                    str_title = elem_artikel_title.Text;

                    IWebElement elem_artikel_prijs =
artikel.FindElement(By.ClassName("promo-price"));
                    str_prijs = elem_artikel_prijs.Text;

                    Console.WriteLine("****** Artikel " + vcount + "
******");
                    Console.WriteLine("Title: " + str_title);
                    Console.WriteLine("Prijs: " + str_prijs);
                    Console.WriteLine("\n");

                }

                vcount++;
            }




        }
    }
}
```

3   LINKS EN BRONNEN

3.1   Youtube

https://youtu.be/Nwai3b2orHA

3.2   Github

https://github.com/jasper-versmissen/CaseStudyWebScraping

3.3   Hoofdbronnen

https://www.youtube.com/watch?v=CpugqTr2j60

https://www.youtube.com/watch?v=gRoMR3NcpPQ&t=37s